[Linguistic Portfolios](https://repository.stcloudstate.edu/stcloud_ling)

Volume 8 Article 2

2019

A Comprehensive Review of F0 and its Various Correlations

Ettien Koffi
St. Cloud State University, enkoffi@stcloudstate.edu

A COMPREHENSIVE REVIEW OF F0 AND ITS VARIOUS CORRELATIONS

ETTIEN KOFFI

ABSTRACT

This paper examines many acoustic characteristics of F0 and their relevance in linguistic analysis. It also highlights correlations between F0 measurements and vowel height, gender, accentedness, and phonation types. The latter is the center piece of the paper. This correlation is needed for a more reliable account of pitch contrasts in accent and tone languages. Two approaches are used in establishing this correlation. The first relies on a subharmonic equation and the second makes use of Critical band calculations. Both yield the same results. The data used to highlight these various correlations come from Peterson and Barney (1952), Hillenbrand et al. (1995), and a fresh set of F0 measurements obtained from 46 speakers of Central Minnesota (17 males and 29 females).

1.0 Introduction

This paper was originally conceived as a handout to answer various questions that students ask in my acoustic phonetics and sociophonetics classes. With each passing semester, more information is added to the handout. What started as a question and answer handout has now morphed into a sizeable publication. It is my hope that students in speech science, acoustic phonetics, and general linguistics will find the insights therein useful in interpreting F0 measurements. One of the biggest challenges in writing a comprehensive review on this topic has been how to organize the flow of information. It has been hard to decide which information to present first, second, third, etc. In spite of my best effort, the structure of the paper has remained a challenge. After several different versions, each with a different structure, I have settled on the current one which consists of two main installments, each with several subdivisions. The first installment is an overview of the main developments in the scholarly study of F0/pitch. The second correlates F0 measurements directly to phonation types.

2.0. Quick Overview of Auditory Pitch Perception Research

Heller (2013:437-528) describes the towering figures who have left an indelible mark on the study of pitch. He devotes 91 pages to these renowned experts. However, for the purposes of this paper, we will highlight Harvey Fletcher (1884-1981) and von Békésy (1899-1972) whose contributions he writes about on pages 507-8. Both are world renowned physicists whose works have revolutionized the contemporary understanding of the auditory perception of F0 and other formant frequencies used in human speech. Fletcher is credited with having invented the modern audiograph machine and the artificial larynx. He held as many as 200 patents. He published a very influential paper in 1940 in which he posited on the basis of mathematical calculations, the frequency response system of the basilar membrane. Allen (1996:1837) summarizes Fletcher's accomplishments as follows:

> The problems that Fletcher and his colleagues studied were so complicated, and took so many years, that it has been difficult to appreciate the magnitude of their accomplishments … In 1918 Fletcher had taken on the toughest problem of all: to quantify and model how we hear and understand speech. This understanding allowed AT&T Bell Labs engineers to develop the necessary specifications of what was to become the largest telephone network

> in the world. It is therefore understandable why his work has had such a great impact on our lives. Almost single-handedly he created the fields of *communication acoustics* and *speech and hearing* as we know them today. Everyone who has ever used the telephone has reaped the benefit provided by this man and his genius. von Bekesy, Davis, Stevens, and Zwicker are some of the names that come to mind when we think of hearing. Bell invented the telephone, and Edison made it into a practical device. Harvey Fletcher may not be as well-known as these men today, but his scientific contributions to the fields of telephony, hearing, and human communication are absolutely unsurpassed.

Yost (2015:49) tops this magnificent tribute by adding the following,

> Fletcher oversaw a litany of psychoacoustic research achievements unmatched in the history of the field, which included measurements of the auditory thresholds (leading to the modern-day audiogram, the gold standard for evaluating hearing loss), intensity discrimination, frequency discrimination, tone-on-tone masking, tone-in-noise masking, the critical band, the phon scale of loudness, and the articulation index. Two of the more important psychoacoustic contributions of the Bell Laboratories years are the critical-band and equal-loudness contours.

Békésy, another physicist, spent nearly two decades doing all sorts of experiments on the ears of cadavers. His findings proved clinically that Fletcher's theory of critical bands was anchored in physiological reality. For his untiring efforts, his scientific genius, and his technological inventiveness in designing instruments to investigate the cochlea and related structures, he was awarded the Noble Prize in Physiology/Medicine in 1961.

Fletcher's and von Békésy's groundbreaking findings have established beyond dispute that the basilar membrane is not only a frequency analyzer but also a transducer of acoustic signals into electrical systems. This tiny organ of 3.2 to 3.5 cm in length contains some 30 critical bands. Each band is 1.3 mm long and specializes in perceiving a specific range of frequencies. The basilar membrane is lined up with more than 12,000 outer hair cells and some 3,500 inner hair cells that act as neurotransmitters. They transmit electrochemical elements to the synapses which, in turn, ferry them to the Central Auditory Nervous System (Yost 2007:285-6). Figure 3 helps to visualize some aspects of the process just described.

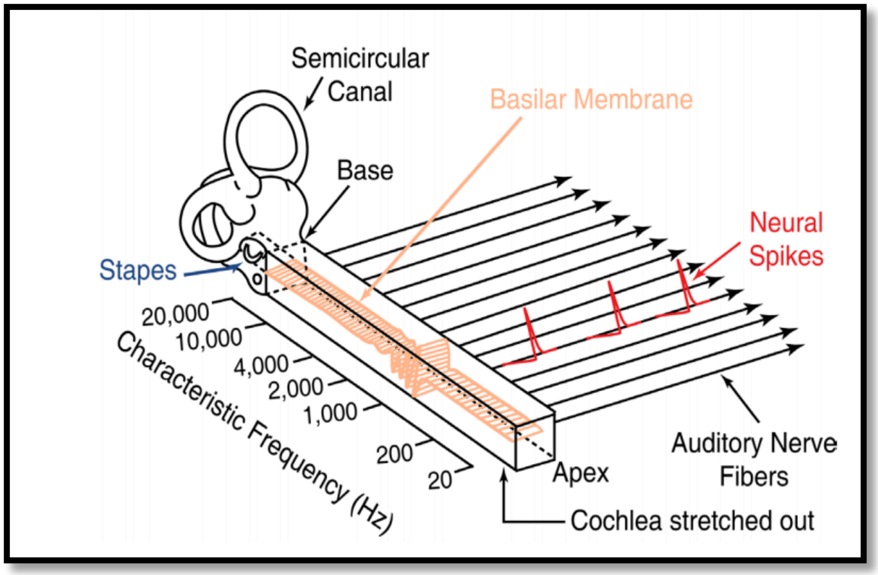
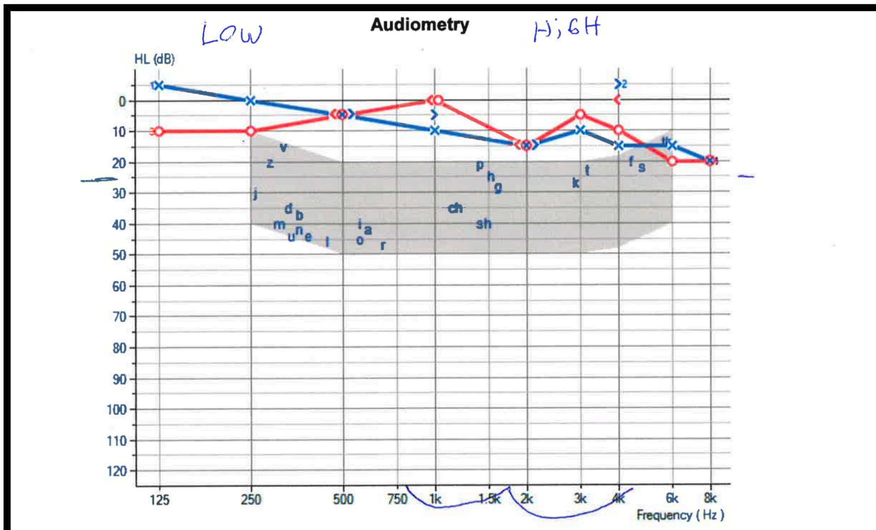


Figure 1: Audibility Range in the Frequency Domain
Sachs, M. B., Bruce, I. C., Miller, R. L., and Young, E. D. (2002). Biological basis of hearing aid design. *Annals of Biomedical Engineering* 30, 157-168. doi:10.1114/1.1458592. Reprinted by permission of © Biomedical Engineering Society.

The portion of the basilar membrane closer to the base perceives the highest frequencies. The frequency-perceiving capabilities of the basilar membrane goes diminishing until it reaches the apex, as shown in Figure 1. Babies and adolescents, for example, can perceive frequencies from 20 to 20,000 Hz. However, auditory acuity goes diminishing with age. The highest frequencies that many normal adults with "perfect" hearing can perceive is between 13,000 to 15,000 Hz range. Audiologists test for hearing acuity in the 8,000 Hz range because these are the frequencies that we need to perceive human speech. Figure 2 shows the results of my pure tone audiometry test results:



Left	125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k
AC	-5	0	5		10		15	10	15	15	20
BC			5		5		15		-5		

Right	125	250	500	750	1k	1.5k	2k	3k	4k	6k	8k
AC	10	10	5		0		15	5	10	20	20
BC			5		0		15		0		

Figure 2: Audiogram Results

Figure 2 is one practical application of Fletcher's and Békésy's seminal research. Their experiments have yielded similar groundbreaking results in a wide variety of scientific and technological pursuits. Suffice it to mention only that critical bands have been endorsed by the American National Standards Institute (ANSI), the International Standardization Organization (ISO), the International Electrotechnical Commission (IEC), and other prestigious world bodies for the manufacturing and engineering of audio products. Critical bands also have important theoretical values. Scharf (1961:215) notes that they are the basic unit of hearing. In other words, if a physical reality is auditorily perceptible, the Critical Band Theory (CBT) can account for it. It goes without saying that CBT is worth knowing and applying to the linguistic research because it can validate or invalidate impressionistic perceptions of pitch.

3.0 Quick Overview of Pitch Production Research

Research in speech production has made great strides thanks to the Source-Filter Theory (SFT) and to technological advancements. SFT was introduced by Gunnar Fant (1960) and expounded on by leading phoneticians such Stevens (2000) and Ladefoged (2006). According to this theory, the larynx is the source, and all the areas above it are the filter. Every millimeter of variation from the source and thereafter affects the quality of the speech sound that we hear. The pitch of speech sounds that humans produce and perceive are affected by four interrelated factors:[1]

1. Length of vocal folds
2. Tension of vocal folds
3. Mass of the vocal folds
4. sub-glottal air pressure

[1] These anatomical variables combine with frequencies in the oral and nasal cavities and contribute a great deal of information that can be used in voice biometric analyses for speaker identification and/or verification. However, in this paper we focus only on F0 and its correlations with phonation types.

The slightest variations in these factors affect F0/pitch. Ammerman (2016:807) likens the vocal folds to plucking the strings of a guitar. The slightest movement affects the quality of the sound that is produced and perceived. We learn from Stevens (2000:5) that the length of the vocal folds in adult females is 1.0 cm, but it is 1.5 cm in adult males. He also notes that the mass of the vocal folds vary from 2 to 3 mm in males and females. In sum, the four factors listed above impact the F0/pitch that humans produce and perceive.

4.0 Laryngoscope Exam of the Larynx and Phonation

Technological innovations have given researchers unprecedented access to the larynx and related structures. Consequently, now more than anytime in human history, we have greater insights into the human speech production mechanism. In the act of speaking, the vocal folds elongate, retract, abduct tightly, or loosely depending on the sounds being produced or the whether or not the speech sounds occur in isolation, in natural speech, or in reading a paragraph. Sophisticated video cameras make it possible to see the movements of the vocal folds live when people are speaking.[2] This gives credence to Stone and Shadle's (2016:54) statements about the hyper-complexity of pitch production:

> It has long been accepted that vocal fold vibration is a complex system. To predict the effect of any physical change accurately, a vocal fold model must include aerodynamic, mechanical, and acoustic elements. We have not yet arrived at comparable models of supraglottal aeroacoustic sources. The study of speech production is an ongoing endeavor. The well-accepted wisdom of any era is subject to revision in the future with the advent of new ideas, new instrumentation, and new research.

Tools such as the one in Figure 3A help researchers and clinicians to have a clear view of a larynx in Figure 3B.[3]

Figure 3A: Laryngoscopy Exam

Figure 3B: Picture of a Normal Larynx

I was blown away by the complexity of speech production and phonation types when I was allowed to observe an otolaryngology examination of female patient related to me. More will be

[2] See the video of a person speaking at https://www.youtube.com/watch?v=y2okeYVclQo. Retrieved on February 19, 2019.

[3] These are not the images of the actual exam. I did not receive permission from the clinician to include his picture in this paper. The pictures in Figure 3A and B are taken from the website of the American Academy of Otolaryngology at https://www.youtube.com/watch?v=2js72BYjZAw. Retrieved on February 19, 2019.

said about it in sections 10.1 to 10.4 where the correlations between F0 and phonation types are discussed.

5.0 A Review of F0s Measurements in English

Peterson and Barney (1952:176-7, 183) were the first to provide a large-scale correlation between F0 measurements and specific vowels in American English. Their study had 76 participants (33 men, 28 women, and 15 children). It was followed 43 years later by Hillenbrand et al. (1995:3099-3100, 3013). Their study had 139 participants (45 men, 48 women, and 46). Excluding the children, the two studies included 154 adults (78 men and 76 women). The average F0s for males in Peterson and Barney and Hillenbrand et al. are respectively 132 Hz, and 130 Hz. The average for the females are respectively 223 Hz and 220 Hz. If we combine the two studies, we can see that the average F0s produced by American English speaking males and females are respectively 131 Hz and 221 Hz.

Words	**heed**	**hid**	**hayed**	**head**	**had**	**hod**	**hawed**	**hoed**	**hood**	**who'd**	**hud**
	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Male (P&B)	136	135	NA	130	127	124	129	NA	137	141	130
Male (Hill)	138	135	129	127	123	123	121	129	133	143	133
Female (P&B)	235	232	NA	223	210	212	216	NA	232	231	221
Female (Hill)	227	224	219	214	215	215	210	217	230	235	218

Table 1: Intrinsic F0 Measurements

Since human beings have the same phonation apparatus, thresholds have been established for studying F0 in all human languages. Fry (1979:68) states that the lowest pitch that a human can produce is 60 Hz. It takes special training and abilities to produce F0s lower than 60 Hz. Similarly, the highest F0 that a human adult can produce when speaking is 500 Hz. Colicky babies can produce F0 beyond 500 Hz. This happens when they are crying. Crying and speaking are two different things! These lower and upper thresholds explain why the default settings in Praat for pitch analysis range from 75-500 Hz (Boersma 2013:9-10).

6.0 Methodology of the Current Study

Table 1 gives us a glimpse of the correlation between F0 and vowel height. More will be said about this in 7.0. As useful as this information is, it fails to provide a much needed correlation between individuals and the phonation type that they use when speaking. In order to uncover such a correlation, a study is undertaken to examine the F0 produced by 46 participants from Central Minnesota (17 males and 29 females). The data was collected over five successive academic semesters. They are a mixture of undergraduate and graduate students who enrolled in my acoustic phonetics and sociolinguistics courses. They fit the prototypical profile of participants in speech science research (Hazan 2017:38). Figure 4 highlights the annotation procedure used to collect the relevant data.

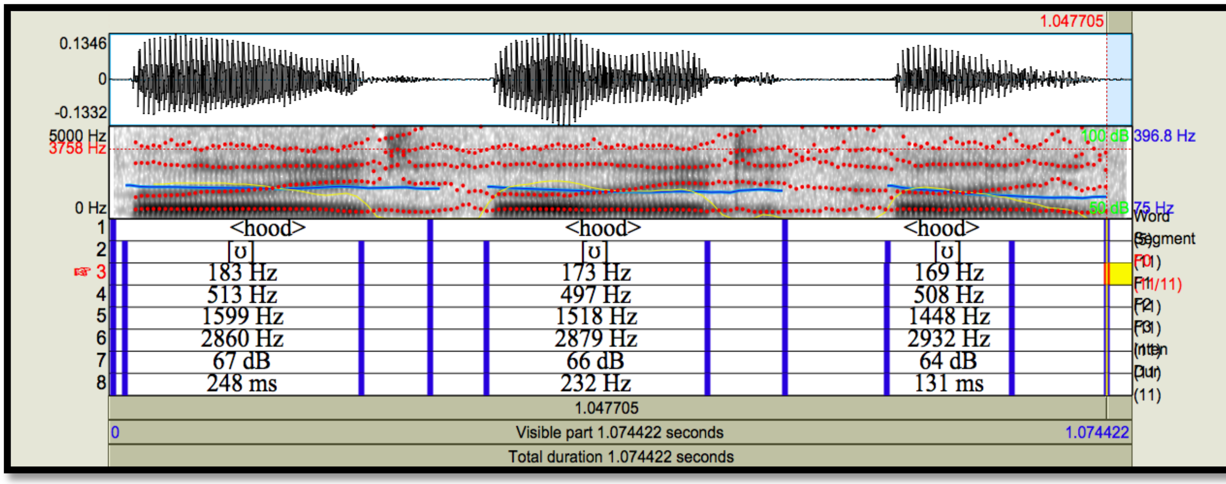


Figure 4: Annotation Procedures

Boundaries were drawn around each vowel and various measurements were taken: F0, F1, F2, F3, Intensity, and duration. Since F0 is our main focus, we will ignore all other correlates. The reported F0 value for each speaker represents the arithmetic mean after three repetitions. The total amount of F0 tokens measured and averaged is 1,518. The 17 male speakers produced 561 tokens (11 x 3 x 17), while the 29 females produced 957 tokens (11 x 3 x 29). They produced the 11 monophthong phonemic vowels of English for this study.[4] Even though other voiced segments can be used to study F0, vowels are researchers' favorite. Esling (2013:122) explains why:

> Vowels, as syllable nuclei, generally longer in duration than consonants, are able to carry many phonotary contrasts. In many languages of the world, they carry phonotary and pitch information, yielding languages with tone contrasts, register contrasts or tonal register contrasts. Breathy-voiced (or whispery-voiced), creaky-voiced and various kinds of harsh-voiced syllables can combine with a range of pitch targets to produce phonological systems with tone along one dimension and phonotary register along another.

6.1 F0 Measurements of Male Speakers

In keeping with Peterson and Barney (1952)[5] and Hillenbrand et al. (1995), the data was not normalized. Watt (2013:91) deems normalization unnecessary when speakers are grouped according to gender. The F0 mean produced by the 17 male speakers of Central Minnesota English (CMNE) is 115 Hz. This is lower by 17 Hz in Peterson and Barney and by 15 Hz in Hillenbrand et al.

[4] The IRB approved the study and all the participants signed an informed consent explaining what the study was about and that they were free to withdraw from the study at any time.
[5] Peterson and Barney and Hillenbrand et al. are abbreviated in Tables as P&B and Hill.

Words	**heed**	**hid**	**hayed**	**head**	**had**	**hod**	**hawed**	**hoed**	**hood**	**who'd**	**hud**
F0 Males	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	114	130	114	122	114	113	114	117	124	126	120
Speaker 2M	146	153	137	144	139	128	140	145	154	162	130
Speaker 3M	102	102	92	98	96	93	92	91	95	93	92
Speaker 4M	148	141	135	117	127	133	126	126	134	130	128
Speaker 5M	105	105	101	98	98	98	99	102	105	106	100
Speaker 6M	118	113	109	109	103	102	96	103	113	126	113
Speaker 7M	110	108	97	104	97	93	98	102	109	108	94
Speaker 8M	123	123	113	117	113	114	116	120	121	123	114
Speaker 9M	123	108	109	106	110	107	103	113	117	119	106
Speaker 10M	122	139	108	119	80	81	82	87	83	82	78
Speaker 11M	123	119	115	115	115	111	110	114	119	117	124
Speaker 12M	98	105	102	100	97	98	103	111	106	109	102
Speaker 13M	113	110	110	113	108	112	111	113	124	124	113
Speaker 14M	153	160	143	137	137	140	139	138	146	145	131
Speaker 15M	133	119	121	111	110	102	115	125	129	139	115
Speaker 16M	158	117	114	115	145	109	109	173	118	136	105
Speaker 17M	129	129	127	128	127	130	127	137	133	137	129
CMNE	**124**	**122**	**114**	**114**	**112**	**109**	**110**	**118**	**119**	**122**	**111**
St. Deviation	17.9	17.3	14.1	12.8	17.5	15.8	15.8	21.1	17.5	19.4	15.1
P&B	**136**	**135**	NA	**130**	**127**	**124**	**129**	NA	**137**	**141**	**130**
Hill et al.	**138**	**135**	**129**	**127**	**123**	**123**	**121**	**129**	**133**	**143**	**133**

Table 2: Intrinsic F0 Measurements for CMNE Males

6.2 F0 measurements of Female Speakers

The mean F0 of female CMNE speakers is 214 Hz, compared with 223 Hz in Peterson and Barney, and 220 Hz in Hillenbrand et al. Here too, we see that CMNE female speakers have a lower F0 than their counterparts in other parts of the USA.

Words	**heed**	**hid**	**hayed**	**head**	**had**	**hod**	**hawed**	**hoed**	**hood**	**who'd**	**hud**
F0 Females	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	186	176	171	179	169	173	169	172	176	187	172
Speaker 2F	224	220	212	222	205	220	211	231	234	269	229
Speaker 3F	277	276	253	245	240	233	226	234	236	253	252
Speaker 4F	248	246	237	234	219	236	227	246	241	255	243
Speaker 5F	191	189	180	180	177	165	172	188	188	218	181
Speaker 6F	199	192	155	190	178	172	170	169	201	199	182
Speaker 7F	243	248	192	251	198	216	209	244	303	301	218
Speaker 8F	217	210	206	204	205	203	200	207	210	224	208
Speaker 9F	225	229	216	225	212	211	209	211	219	221	217
Speaker 10F	241	236	240	230	230	227	231	232	241	248	236
Speaker 11F	163	191	192	192	190	172	190	185	198	174	181
Speaker 12F	235	223	233	205	201	208	211	235	232	248	203
Speaker 13F	207	196	200	192	194	189	190	197	196	219	191
Speaker 14F	207	191	189	190	186	190	191	193	193	203	173
Speaker 15F	223	211	227	221	219	220	221	215	229	224	227
Speaker 16F	219	226	227	230	238	231	228	225	220	213	234
Speaker 17F	217	224	208	211	205	239	187	199	216	233	196
Speaker 18F	266	282	270	244	237	254	245	259	253	269	249

Speaker 19F	207	214	211	349	216	205	211	214	209	203	225
Speaker 20F	245	232	236	222	207	232	220	213	217	224	217
Speaker 21F	242	246	248	234	243	233	254	225	234	245	240
Speaker 22F	236	218	210	205	209	190	211	208	216	225	202
Speaker 23F	223	258	207	244	238	220	234	236	241	249	229
Speaker 24F	215	222	221	209	212	212	205	206	212	237	214
Speaker 25F	210	214	188	191	169	173	182	179	185	167	167
Speaker 26F	226	234	223	223	201	208	217	212	215	220	204
Speaker 27F	177	160	154	157	169	153	161	159	178	184	182
Speaker 28F	224	236	224	229	212	204	211	221	229	238	170
Speaker 29F	207	215	212	216	203	186	195	206	207	222	202
CMNE	**220**	**221**	**211**	**218**	**206**	**206**	**206**	**211**	**218**	**226**	**208**
St. Dev.	24.7	27.6	27.4	34	21.8	25.6	23	24.3	25.9	29.6	25.4
P&B	**235**	**232**	NA	**223**	**210**	**212**	**216**	NA	**232**	**231**	**221**
Hill et al.	**227**	**224**	**219**	**214**	**215**	**215**	**210**	**217**	**230**	**235**	**218**

Table 3: Intrinsic F0 Measurements for CMNE Females

7.0 Correlation between F0 and Vowel Height

Table 4 contains the F0 measurements from Peterson and Barney, Hillenbrand et al., and the data from CMNE. All three sets of data concur that there is a correlation between vowel height and their intrinsic F0 measurements. High vowels consistently have higher F0 values than non-high vowels in all three dialects of American English.

Words	**heed**	**hid**	**hayed**	**head**	**had**	**hod**	**hawed**	**hoed**	**hood**	**who'd**	**hud**
	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Male (P&B)	136	135	NA	130	127	124	129	NA	137	141	130
Male (Hill et al.)	138	135	129	127	123	123	121	129	133	143	133
CMNE Male	124	122	114	114	112	109	110	118	119	122	111
Aggregated F0	**132**	**130**	**121**	**123**	**120**	**118**	**120**	**123**	**129**	**135**	**124**
Female (P&B)	235	232	NA	223	210	212	216	NA	232	231	221
Female (Hill et al.)	227	224	219	214	215	215	210	217	230	235	218
CMNE Female	220	224	211	218	206	206	206	211	218	226	208
Aggregated F0	**227**	**226**	**215**	**218**	**210**	**211**	**210**	**214**	**226**	**230**	**215**

Table 4: Vowels with Same F0 Measurements

The correlation between F0 and vowel height is also widely attested in other languages. Ohala (1978:29) writes that "It has been noted over 50 years that, other things being equal, the average pitch of vowels shows a systematic correlation with height, that is, the higher the vowel, the higher the pitch.... The difference between high and low vowels may be as much as 25 Hz." Crothers (1978:115) contends that the correlation is universally attested. Lehiste (1973:70-1) offers a physiological explanation for why this may be the case:

> There appears to be a physiological reason for the fact that high vowels are associated with a relatively high fundamental frequency. As mentioned above, fundamental frequency increases with either increased rate of airflow or increased tension of the vocal folds (or a combination of the two). In the articulation of high vowels, the tongue is raised towards the roof of the mouth. Now, the muscles constituting the tongue are attached to the superior part of the hyoid bone and some of the laryngeal muscles are attached to the inferior part. When the tongue is raised, the larynx tends to be pulled upwards and the laryngeal muscles

are stretched. This increases the tension of the vocal folds and causes the increase rate of vibration (Lehiste 1973:70-1).

In addition to high vowels, coronals, that is, [θ, ð, t, d, s, z, ʃ, ʒ, tʃ, dʒ, l, r], have the same effect on the larynx. The feature [+stiff] has been used to describe some of these segments. Stevens (2000:72) notes that "an increase in stiffness gives rise to an increase in frequency." The observed increase in frequency ranges from 10 to 30% (Stevens 2002:470-7). The increase is most significant when a consonant that has the feature [+stiff] is followed immediately by [i] or [u]. Most notable is the case of the voiceless aspirated stops, [t] and [k] in syllable onsets (Hombert 1978:87). Stevens (2000:535, 538) includes [ɹ] among the segments that have an augmentative effect on F0. However, Jones (2013:144) advises caution about correlating F0 with specific segments, noting that "uncertainty surrounds the mechanism underlying the widespread phenomenon of intrinsic F0, in which high vowels have a higher F0 than low vowels, other things being equal." Hombert (1978:91) too is skeptical, pointing out that there are cases when low vowels have higher F0 than high vowels. Even though these researchers have expressed skepticism, they have not produced actual data to support their contrarian claims.

8.0 Correlation between F0 and Biological Gender

The anatomical differences in the larynx of adult males and females were briefly mentioned in 3.0. They are largely responsible for differences in the pitch in male and female speech. On average, male F0 is 120 Hz in Fry (1973:68)'s study, and 133 Hz in Miller (1989:2122)'s study. For females, they both have 225 Hz. Stevens (1998:1232) gives the ranges of 80-160 Hz for men and 170-340 Hz for women. These values are consistent with those listed in Table 5:

Words	**heed**	**hid**	**hayed**	**head**	**had**	**hod**	**hawed**	**hoed**	**hood**	**who'd**	**hud**
	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Male (P&B)	136	135	NA	130	127	124	129	NA	137	141	130
Male (Hill et al.)	138	135	129	127	123	123	121	129	133	143	133
CMNE Male	124	122	114	114	112	109	110	118	119	122	111
Aggregated F0	**132**	**130**	**121**	**123**	**120**	**118**	**120**	**123**	**129**	**135**	**124**
St. Deviation	7.57	7.50	10.60	8.50	7.76	8.38	9.53	7.77	9.45	11.59	11.93
Female (P&B)	235	232	NA	223	210	212	216	NA	232	231	221
Female (Hill et al.)	227	224	219	214	215	215	210	217	230	235	218
CMNE Female	220	224	211	218	206	206	206	211	218	226	208
Aggregated F0	**227**	**226**	**215**	**218**	**210**	**211**	**210**	**214**	**226**	**230**	**215**
St. Deviation	75.2	75.1	5.6	72.4	69.6	70.1	69.5	4.2	74.9	75.2	70.4

Table 5: F0 Measurements by Gender

In general, female F0s are 57% higher than males'. However, to make calculations easier, 50% is taken as the default. Another way of calculating female F0s, where such data is not readily available, is to multiply male F0 values by 1.7 (Kent and Read 2003:191).

9.0 Correlation between F0 Regional Accents

Tone languages use F0 information to contrast lexical and/or grammatical meaning. However, accent languages don't. Since English is an accent language. The following statement by Miller (1989:2128) must be understood within this context, "…Under most conditions, the identity of a perceived vowel depends strongly on the formant values of the spectrum and is independent of voice pitch." Thus, the main function of F0 in English is paralinguistics, namely providing information about grammatical mood or even the psychological mood of the speaker. However, when we consider the data in Table 6, one is tempted to say that F0 has regional indexical values, that is, it is a marker of accentedness.

Words	heed	hid	hayed	head	had	hod	hawed	hoed	hood	who'd	hud
	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Male (P&B)	136	135	NA	130	127	124	129	NA	137	141	130
Male (Hill et al.)	138	135	129	127	123	123	121	129	133	143	133
CMNE Male	124	122	114	114	112	109	110	118	119	122	111
Female (P&B)	235	232	NA	223	210	212	216	NA	232	231	221
Female (Hill et al.)	227	224	219	214	215	215	210	217	230	235	218
CMNE Female	220	224	211	218	206	206	206	211	218	226	208

Table 6: F0s and Regional Dialects

In sociophonetics circles, F2 is seen as the main or even the only acoustic correlate that carries accentedness (Kent and Read 2002:111, Koffi 2018:80-94). Yet, it is hard to ignore that the substantial differences in F0 between CMNE, Peterson and Barney's the General American English (GAE) data and the Midwest data from Hillenbrand et al. A quick glance at the data shows substantial differences in F0 between the three dialects. The overall F0 mean in GAE is 132 Hz in male speech, 130 Hz in the Midwest, and 115 Hz in Central Minnesota English (CMNE). There is a difference of 17 Hz between GAE and CMNE speakers, 15 Hz between CMNE and Midwest speakers. In female speech, we see the same large differences. In GAE, the F0 is 223 Hz, 220 Hz in the Midwest, and 214 Hz in CMNE. The differences are respectively 9 Hz between CMNE and GAE and 7 Hz between CMNE and Midwest. Furthermore, we learn from Young (2011:609), Lehiste (1970:64), Gandour (1978:57), and Stevens (1998:228) that the human ear can detect pitch variations as little as little as 0.3% (rounded up to 1 Hz). Given this auditory hypersensitivity to F0, it is hard to imagine that F0 does not play a significant role in dialect variation.

This is an aside, but an aside worth making, namely that the correlation works only with [+high] vs. [-high] vowels. It does not work among high vowels, nor does it discriminate among [-high] vowels. A case in point is [ʊ] vs. [ʌ] (133 Hz) which have the same F0 in Hillenbrand et al. in male speech. In CMNE, males produced [e] (114 Hz) and [ɛ] identically, while females did the same for [æ] (206 Hz) vs. [ɔ], and [ɔ] (206 Hz) vs. [ɑ]. Among high vowels, we see that in Peterson and Barney, the F0 of [ʊ] (232 Hz) is higher than [u] (231 Hz) in male speech.

10.0 An Overview of Phonation Types

Before embarking on the study of the correlation between F0 and phonation types, a brief overview of phonation types is in order. A plethora of terms exist to describe the different types of voices that people use to speak. Labels such as vibrato, flutter, rough, hoarse, whispery, harsh, and the like are found in the literature. However, Titze (1994:8) notes that "Such terms have no mathematical definitions. … no numbers or physical units of measurement need to be attached to them, although some of them can be rated psychophysically." Gerratt and Kreiman (2001:366) opine that the proliferation of terms is rooted in the diversity of academic disciplines (music, linguistics, speech pathology, otolaryngology) that work with voice. They approach phonation with their own sets of labels. Being a linguist, I will focus on the four phonation types that are commonly discussed in my discipline:

1. modal voice
2. breathy voice
3. creaky voice
4. falsetto voice.

It should be noted right away that it is an overgeneralization to claim that there are only four phonation types because, in reality, the folds can a assume many different positions in the act of speaking. Berry's (2001) experimental simulations yield at least 30 different configurations. The above-mentioned subcategorizations refer to a person's overall voice quality. With this caveat in mind, laryngoscopic images will be used to illustrate each phonation type. A family member was seen by an Ear Nose and Throat doctor and a clinician. I was allowed to observe the procedures and obtained the illustrative images that follow. Permission was granted to me to use the images in this paper. The images below represent various states of larynx, glottis and related structures when the clinician asked the patient to pronounce the vowel sound [i] in "fleece" in multiple ways. In otolaryngology exams, [i] is the preferred vowel because it helps evaluate the health of the vocal folds. A very sophisticated, state-of-the art laryngoscope such as the one in Figure 3A, but with a more powerful resolution (x 800) was inserted in the nostril of the female patient. The clinician asked her to imitate him as he produced [i]s in various which match the aforementioned phonation types.

10.1 Modal Voice Phonation

A modal voice is produced when the vocal folds are only slightly abducted. This is the most common phonation type. It is the default voice that most speakers use when they are talking normally. This phonation type is unmarked, meaning that, if a speaker uses it, the hearer's attention is not drawn to it. Figure 5 displays a laryngoscopic image of the patient imitating the clinician after he produced a modal [i].

Figure 6: Modal Phonation of [i]

We know that her [i] is modal because its F0 of 231 Hz is on par with other F0 in Table 6. Additional pieces of evidence are provided in 11.2 that supports this conclusion (see Table 8).

10.2 Breathy Voice Phonation

Breathy voice is obtained when the vocal folds are further apart, air molecules from the lungs gush through the glottis freely. Catford (2001:207) describes it as follows, "The glottis is wide open with high velocity airflow so that the vocal folds are flapping in the breeze." Gerratt and Kreiman (2001:377-8) differentiates between "extremely breathy," "slightly breathy," and other types of breathiness. In Figure 7, the patient imitates a breathy [i] produced by the male clinician. In so doing, she produced an exaggerated breathy [i].

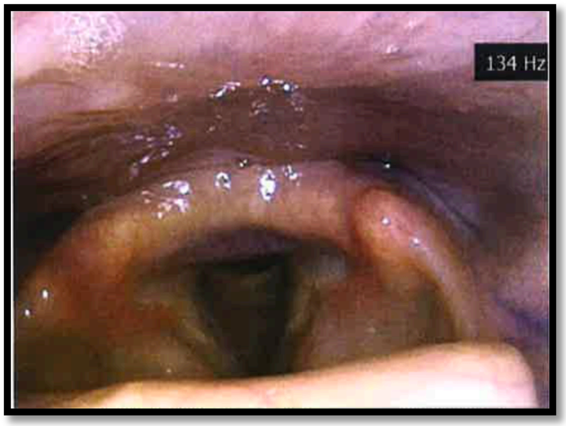


Figure 7: Breathy Phonation of [i]

For the most part, females are poor imitators of male voice. We see here that the patient cannot lower her vocal folds so as to produce the breathy [i] produced by the male clinician. An F0 of 134 Hz is too low for most women to produce.

10.3 Creak Voice Phonation

The creaky voice is also known as "vocal fry." In producing this voice quality, Catford (2001:207) writes that "The glottis is closed along most of its length but with a very small vibrating segment near the front end through which low-frequency bursts of air escape." Ladefoged and Maddieson (1996:53) provide a physiological explanation and description, "The arytenoid cartilages are much closer together than in modal voice. Creaky voice also involves a great deal of tension in the intrinsic laryngeal musculature, so that the vocal folds no longer vibrate as a whole. Sometimes the parts of the vocal folds close to the arytenoid are held too tightly together to be able to vibrate at all; on other occasions the ligamental and arytenoid parts vibrate separately, so that they are out of phase with one another." Keating et al. (2015) add that there are in fact several types of creaky voice depending on where the closure occurs and where the opening is.

The clinician did not produce a creaky [i] for the patient to imitate. However, Figure 8 shows a very good picture of what creaky voice phonation looks like:[6]

[6] YouTube image https://www.youtube.com/watch?v=y2okeYVclQo. Retrieved on February 19, 2019.

Figure 8: Creaky Voice/Vocal Fry

Gerratt and Kreiman (2001:376) indicate that the participants in their study identified creaky voice/vocal fry with more than 95% accuracy. Casual observations indicate creaks occur most frequently when the voice is trailing towards a pause. This phonation type has received a lot of press lately because Millennials and Generation Z speakers have adopted it as their stylish way of speaking. More and more commercials and political ads are featuring it. In advertisement, the creaks are deliberately overdone to draw attention to the person selling a product. For now, public opinion about creaky voice/vocal fry is overwhelmingly negative. However, this is expected to change as this phonation type spreads across generations of women. It is already making inroads among some male speakers.

10.4 Falsetto Voice Phonation

Falsetto is a phonation type that is not ordinarily used. It surfaces in baby-talk, foreigner talk, or in talking to pets. Occasionally, it is used when talking to the elderly. Recently, I was visiting somebody in a hospital and overheard a nurse using this voice quality when interacting with a patient. In the TV show Big Bang Theory, Bernadette speaks in a falsetto voice all the time. When using this voice quality, the speaker elongates his/her vocal folds. In Figure 8, the patient imitates a falsetto [i] produced by the clinician.

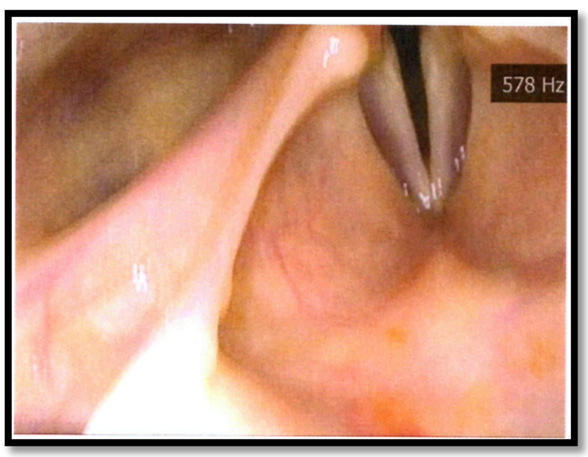


Figure 9: Falsetto Voice

It can be seen very clearly that a falsetto [i] is used because the vocal folds are more elongated than usual and the supporting structures are also stretched. The opening in the glottis is very small. As a result, a high pitched [i] with an F0 578 Hz is produced.

11.0 Correlation between F0 and Phonation Types

One of the unanswered questions in acoustic phonetics is the following: Is there a correlation between F0 values and phonation types? Impressionistically, people make such correlations all the times. Ladefoged (2003:27) remarks that "The ultimate authority in all phonetic questions is the human ear." Gerratt and Kreiman (2001:366) add that "Listeners can easily judge similarities and differences among auditory stimuli and such judgments provide empirical evidence for the existence of distinct vocal phenomena. Once these are determined, acoustic and physiological correlates can be identified and used to confirm or even identify the classification of an utterance." For these reasons, researchers anticipate that such a correlation should exist. The search has gone on unabated for a while. Various schemes have been tried to account for it. Watt (2013:97) mentions the following: 1) jitter ratios in the voice, 2) spectral tilt, 3) F1, 4) F2, 5) duration, 6) intensity, 7) H1*-H2* differences. Keating et al. (2011:1047) discuss many more and ingenious approaches to uncover the said correlation. So far, the golden methodology has remained elusive. In the following section, I introduce two methods. They yield the same results. The first approach is based on a subharmonic frequency equation, while the second is anchored in critical band calculations.

11.1 Subharmonic Frequency Equation and Phonation Types

Gordon and Ladefoged (2001:397) quote Stevens as saying that "Phonation differences can be quantified through a number of phonetic measurements, even if certain physiological or auditory properties defining these types are harder to define". In other words, measurements may hold the key to uncovering the correlation between F0 and phonation types. The measurements used in this regard are based on the following subharmonic frequency equation:

$$\text{Subharmonic Frequency} = \frac{\text{Mean Frequency}}{2 \text{ or } 3.4}$$

This equation is by no means new. We find it in Titze (1994:10). It is also discussed and exemplified in Lehiste (1970:59). Before using the equation, an explanation for why the denominator has two sets of numbers is in order. Subharmonic frequency values are the same for all adult males and females (Gerratt and Kreiman 2001:376). The ratio of 2 is used when one is calculating subharmonic frequencies for men, while the ratio of 3.4 is used for women. The latter takes into account the fact that females' F0s are 1.7 times higher than males (see Kent and Read 2003:191). Thus, the ratio of 3.4 is obtained by multiplying 1.7 by 2. With this, we can now calculate the subharmonic frequency of any F0 if we know the gender of the speaker.[7] Lehiste (1970:59) shows that the subharmonic frequency of 150 Hz in male speech is 75 Hz. If a woman produced a segment with the same F0 value (this is rather unlikely in most cases), the subharmonic frequency would be 44 Hz. A correlation between F0 measurements and phonation types can be established this way. We know from Lehiste (1970:59) and Keating and Kuo (2012:1055) that the subharmonic frequency of 40 Hz corresponds to creaky voice. This threshold is known and widely accepted. Starting from it and moving up by increments of 10 Hz, we can very easily correlate F0 and phonation type by referring to the ranges in Table 6:

[7] Campbell (1997:1442) correlates subglottal pressure with voice biometrics and speaker recognition.

N0	Voice Quality	Subharmonic Frequencies
1.	Creaky	≤40 Hz
2.	Breathy	41to 50 Hz
3.	Modal	51to 60 Hz
4.	Falsetto	≥70 Hz

Table 6: Subharmonic Frequency and Voice Quality

These correlations work very well except in cases involving "slightly breathy" voices.[8] On occasion, they can be mistaken for modal voice. This is so because, as Gerratt and Kreiman (2001:377-8) have noted, hearers do not often agree with phonation types that are borderline between two auditory categories:

> There is a continuum of breathiness: "extremely breathy, slightly breathy, and so on. In fact, listeners do not even seem to agree whether a voice is or is not breathy, except in cases where the voice is nearly aphonic… Thus, breathiness seems to differ from the other kinds of phonation discussed in this paper, in that it apparently does not form a coherent perceptual category, and it varies continuously rather than categorically from modal phonation, … Thus, in the absence of a phonological contrast (i.e., a difference in meaning), categorically separating breathy from modal phonation is impossible, because there does not appear to be a single physiological or acoustic cue, or even a combination of cues, that consistently and reliable indicates [that] breathiness.

This is a well-placed caveat. Yet, I'm confident that the correspondences proposed in Table 6 are a step in the right direction. It is worth noting that there are little "flaws" in the measurements of nearly all physical events. Take the measurements and categorization of tornadoes as an example. The National Oceanic and Atmospheric Administration (NOAA) still relies on Fujita rating system despite its "flaws."[9] NOAA adds the following caveats to the classification of tornadic events:

1) Nobody knows the "true" wind speeds at ground level in most tornadoes.
2) The amount of wind needed to do similar-looking damage can vary greatly, even from block to block or building to building.

The fact that there may be slight overlapping between breathy and modal voices does not invalidate the information in Table 6 above, nor the one in Tables 7 and 8 below.

[8] Lehiste (1970:59) reports that creaky voice can have subharmonic frequencies that range from 23 to 73 Hz. Lindblom (2009:3) seems to suggest that that males produced creaky voice at 75 Hz and females at 100 Hz. There are reasons to believe that some reporting errors occurred in his text. There is a widespread consensus that creaky voice is not produced with such high subharmonic frequencies. Gerratt and Kreiman (2001:375-6) also mention having vocal fry as low as 7 Hz and as high as 78 Hz. Most likely 7 Hz is a typo, otherwise it would be the lowest fry ever encountered. We take 40 Hz as the JND for creakiness even though some creaky voices may have higher subharmonic frequencies than 40 Hz. However, we do not expect creaks to be as high as 50 Hz. If they are, they qualify as "extremely breathy," to borrow Gerratt and Kreiman's terminology.

[9] Source: https://www.weather.gov/mkx/taw-tornado_classification_safety. Retrieved on September 25, 2018.

11.2 Critical Bands and Phonation Types

If people can hear a difference between two phonation types (refer back to the quote from Gerratt and Kreiman 2001:366 and Ladefoged 2003:27 in 11.0), then such auditory perceptions can be accounted for by the Critical Band Theory (CBT) because, according to Scharf (1961:215), "critical bands are the basic units of hearing." The groundwork for doing so was laid in Koffi (2016, 2018). The same logic is used here to establish correlations between F0 measurements and phonation types. The correspondences are displayed in Tables 7 and 8:

N0	Voice Quality	Lower Limits	Center Frequency	Upper Limits
1.	Creaky	71	80	88
2.	Breathy	89	100	113
3.	Modal	114	125	141
4.	Falsetto	142	160	176

Table 7: Critical Bands and Voice Quality for Men

N0	Voice Quality	Lower Limits	Center Frequency	Upper Limits
1.	Creaky[10]	139	157	177
2.	Breathy	178	196	219
3.	Modal	220	251	282
4.	Falsetto	283	315	353

Table 8: Critical Bands and Voice Quality for Women

Adjustments have been made to 1/3 frequency responses in Table 7 for Table 8 because of the anatomical differences between males and females discussed in 3.0, 8.0, and in 11.0.

I'm confident that these correlations work but it is difficult to find data for verification. Most studies do not report raw F0 measurements of individual speakers. Publications often report their statistical analyses without providing the F0 raw measurements which lead to their statistical conclusions. In spite of my best efforts to find data to verify the correlations in Tables 7 and 8, only one study can be used to do so. Keating and Kuo (2012:1053, Table III, p. 1056) measured the F0 produced by 23 English and 23 Mandarin speakers. Each language group contained 11 men and 12 women. Kuang (2017:1701) interprets the range of 77-85 Hz in male speech to be creaky voice. The range of 136-150 Hz in female speech is also interpreted as creaky voice. Note that their findings agree completely with the correlations in Tables 7 and 8. The more raw F0 measurement data become available, the more confident researchers will feel using CBT to correlate F0 and phonation types. While waiting for independent data to accumulate, let's apply these insights to the data produced by 46 speakers from Central Minnesota. Their F0 measurements were displayed earlier in Tables 2 and 3.

12.0 Application of CBT to Phonation Types in the CMNE Data

The overall F0 mean that the 17 male speakers produced is 116 Hz. This corresponds to a modal voice phonation type in Table 7. When we scrutinize that data further, we see that the male participants produced 7 of their 11 vowels with a modal voice quality. This amounts to 63.63%.

[10] We begin the correlation between F0 and phonation with the low pitch measurements because, except for voice abnormalities, extra low pitch is rare among women (see Koffi 2018:126).

The vowels [æ, ɑ, ɔ, ʌ] were produced with a breathy or creaky voice quality. When we dig deeper into the data and examine the pronunciation of individual speakers, we see that 5 out the 17 male speakers (29.41%) of the participants have a breathy voice quality. These are Speakers 5M, 6M, 7M, 9M, and 12M. Speaker 13M is borderline between modal and creakiness. Two participants, Speakers 3M and 10M have creaky voice (11.76%). Overall, 7 out 17 speakers (41.17%) have a non-modal voice.

As for the female participants, overall, they produced their vowels with a breathy voice quality. The F0 mean of their vowels is 214 Hz. Only three vowels, [i, ɪ, u], are produced with a modal voice quality. The remainder of the vowels, [e, ɛ, æ, ɑ, ɔ, o, ʊ, ʌ], 72.2%, have a breathy or creaky voice quality. There is interspeaker variability. Speaker 1F and 27F produced their vowels with a creaky voice. The rest of the female participants are almost evenly split between breathy voice (14 speakers) and modal voice (13 speakers). Percentage-wise, 44.82% of the females produced their vowels with breathy voice compared with 29.41% of males. This finding agrees with others summarized in Esposito (2010:182), namely, "Acoustic analyses showed that on average American-English–speaking females were breathier than males."

13.0 Spectrographic Analyses of Voice Quality

A disproportionate amount of attention has been paid to inspecting spectrographs in phonation studies. The spectrographs below are used to highlight the advantages and disadvantages of relying almost exclusively on spectrographic evidence. I agree in principle with Gerratt and Kreiman (2001:366) that "Although two voice samples may have been produced rather differently, and the acoustic waveforms may look rather different, these differences are important only if they result in a perceptually salient difference in vocal quality." The annotated spectrograph below depicts three aspects of my pronunciation of [ɛ] in <heck> (see Koffi 2012:1-4):

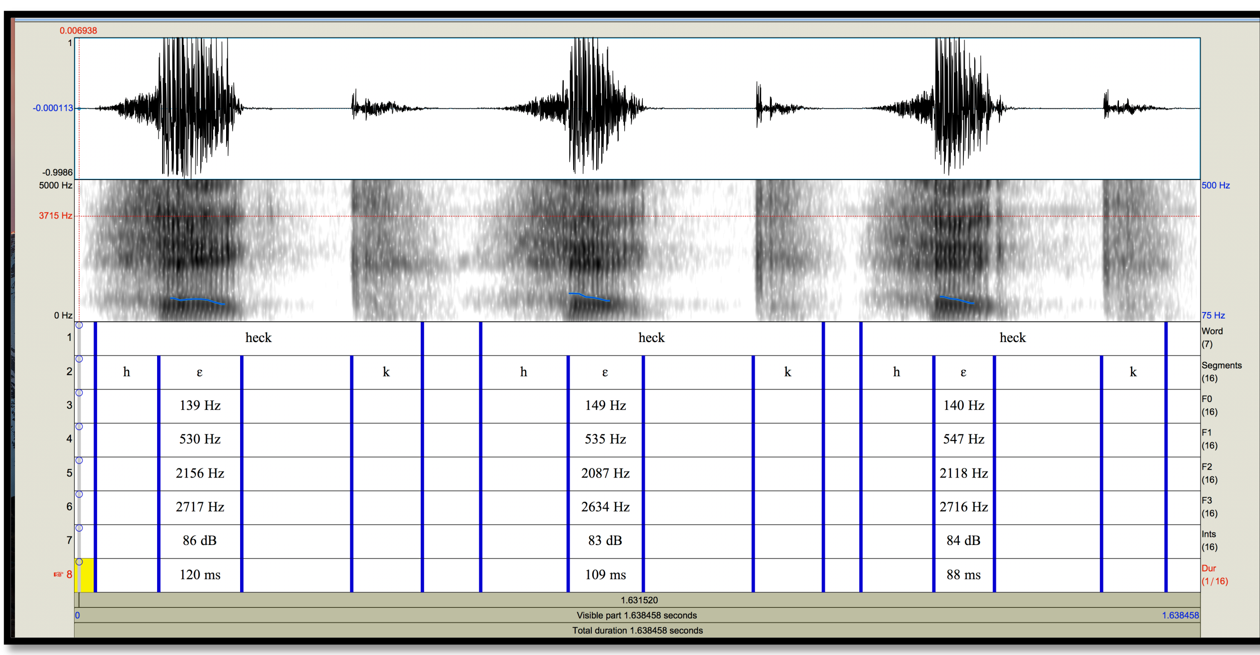


Figure 9: Spectrographs

A casual inspection reveals that each of the three waveforms are slightly different. Furthermore, in the second and third iterations of [ɛ], my vocal folds do not vibrate the whole way through the vowel, as shown by the pitch track. I zoom in on the last iteration and display it in Figure 10 for a better look:

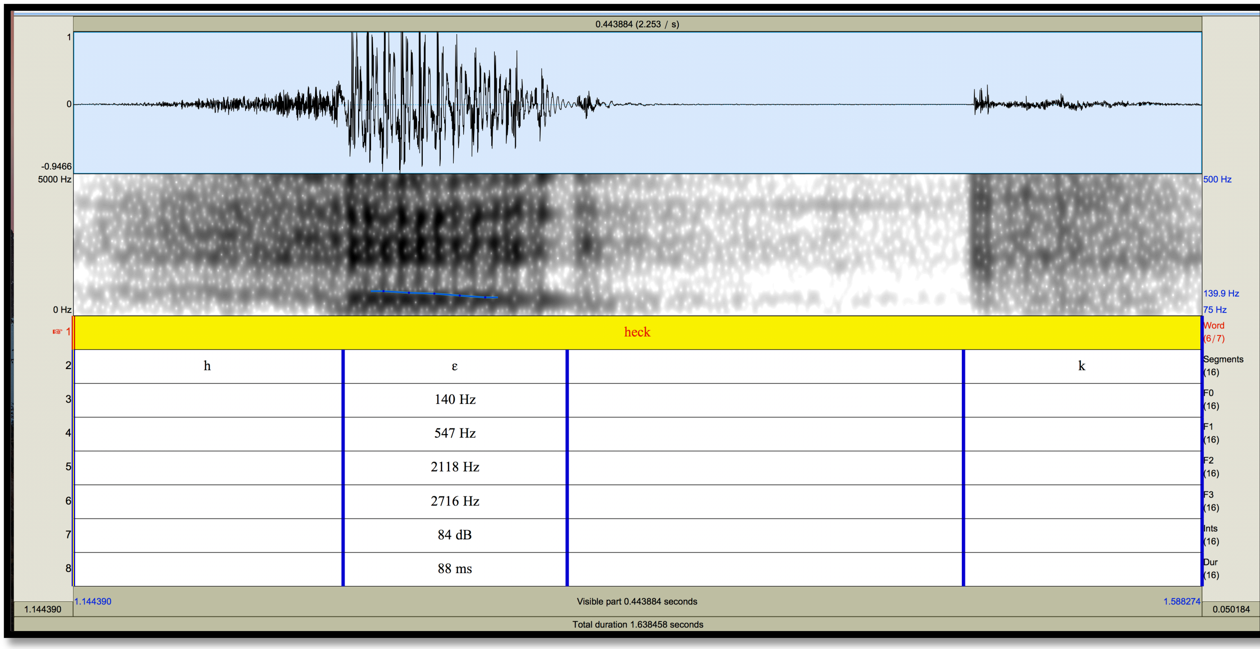


Figure 10: Zoomed in Spectrograph

We see that a good portion of the final edge of the vowel is completely devoiced or voiceless. This is an indication of creakiness or breathiness even though the overall voice quality perceived by the hearer is a modal voice. Gordon and Ladefoged (2001: 393) found a similar situation in their study. They commented on it saying,

> Confinement of nonmodal phonation to portions of vowels is not only a feature of languages that use tone contrastively. In Hupa, which does not have contrastive tone, creaky voice and breathy voice spread from syllable-final ejectives and voiceless obstruents, respectively, onto the latter half of a preceding long vowel and not onto short vowels at all.

The quote is about Hupa but it has insight for other languages, namely tone languages. The first sentence of the quote indicates that nonmodal phonation in portions of vowel is common in tone languages. Coincidentally, Anyi, my native language is a tone language. This would explain why I produce partially voiced vowels. There have been occasions when I produce voiceless vowels! Ladefoged and Maddieson (1996:315) also note that voiceless vowels occur in the Bantu languages of the Congo basin. The take-away from the spectrographic analysis of phonation types is that it reveals more information than a speaker's ears can perceive. This may be the reason why some researchers have found more types of creaky voices or breathy voices than can be possibly perceived by the naked ears. It would be very useful going forward if researchers would display at least three iterations of the same segment for visual inspection. Simply choosing a pronunciation that fits the description that one intends to highlight does not tell the whole story.

14.0 Linguistic Uses of Phonation Types

Human beings are adept at using subtle variations in phonation types to convey many types of subtle paralinguistic information such as the ones that we have been discussing thus far. For example, Gordon and Ladefoged (2001: 391) write that "Nonmodal phonation, especially creaky voice, is commonly used cross-linguistically as a marker of prosodic boundaries, either initially and/or finally." However, in general, phonation types are not phonemically contrastive, except for in some isolated languages. Ladefoged and Maddieson (1996:57, 82) report that in Indo-Aryan languages such as Hindi and Marathi, breathy voice quality is used phonemically to discriminate among stop consonants. As for African languages, on page 87, they note Lendu contrasts creaky voice with modal voice in bilabial, alveolar, and velar stops. Mention is also made of Somali, Hausa, and some Chadic languages where creaky voice is used contrastively with modal voice in the production of stop consonants (Ladefoged and Maddieson 1996:53). Furthermore, they report that "breathy voiced stops also occur in a number of African languages, including Owerri Igbo (p. 62). Finally, Ladefoged and Maddieson (1996:317) mention Jalapa Mazatec as a language that has a three-way phonemic contrast between modal, breathy, and creaky vowels. Considering that there are some 6,000 languages in the world today, we can reasonably conclude that the paralinguistic usages of phonation types are far more important than their linguistic ones.

15.0 Summary

This study has done exactly what it purposed to do, i.e., provide a comprehensive review of F0. The review has summarized, albeit at a very fast pace, salient aspects of more than 100 years of auditory perception research of F0/pitch. A survey of the articulatory and physiological components of pitch production was also provided, including a brief discussion and illustration of state of the art advances and uses of laryngoscope imaging techniques to visualize phonation. A large set of data from Peterson and Barney (1952), Hillenbrand et al. (1995), and from the 46 speakers of Central Minnesota English helped establish various correlations, including the ones between F0 and gender, F0 and accentedness, and F0 and vowel height. More importantly, demonstrations were made of how F0 measurements could be correlated directly with phonation types. This called for appealing to a subharmonic algorithm and to critical band calculations. Both yield the same results but preference was given to critical band-based analysis because it has a long track record of being successfully applied to acoustic research and to the manufacturing of audio product and sound level meters. The demonstration was based on data from 46 speakers of Central Minnesota English. The preliminary results suggest that F0 measurements can be correlated directly with phonation types. However, large scale data is much needed for independent confirmation. It would be helpful if future researchers provide F0 measurements of individual participants in their studies instead of reporting only the aggregated mean of all the participants. Such large scale data will help determine if the information in Tables 7 and 8 predict accurately the phonation types used by the participants in their studies.

ABOUT THE AUTHOR

Ettien Koffi, Ph.D. in linguistics from Indiana University, teaches linguistics at Saint Cloud State University, MN. Author of many peer-reviewed articles on various topics in linguistics and of four books: *Language and Society in Biblical Times* (1996), *Paradigm Shift in Language Planning and*

Policy: Game Theoretic Solutions (2012), *Applied English Syntax* (2010, 2015), and the *New Testament in Anyi Morofu* (2017), a task which took over 25 years. Specializing in acoustic phonetics, dialect variation, and emergent orthographies, his current research centers on speech acoustics of L2 English (within the Speech Intelligibility Framework), Central Minnesota English, and Anyi. He can be reached at enkoffi@stcloudstate.edu.

Reference

Allen, Jont B. 1996. Harvey Fletcher's Role in the Creation of Communication Acoustics. *Journal of the Acoustical Society if America* 99 (4):1825-1839.

Berry, David A. Mechanisms of Modal and Nonmodal Phonation. *Journal of Phonetics* 29:431-450.

Boersma, Paul. 2013. Acoustic Analysis. In *Research Methods in Linguistics* ed. by Robert Podesva and Devyani Sharma, pp. 1-21. New York: Cambridge University Press.

Boersma, Paul & Weenink, David (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.17, retrieved 21 April 2016 from http://www.praat.org/
Campell, Joseph P. (1997:1452-3). Speaker Recognition: A Tutorial. *Proceedings of the*

Campbell, Joseph P. Speaker Recognition: A Tutorial. *Proceedings of the IEEE* 85 (9): 1437-1462.

Catford, J. C. 2001. *A Practical Introduction to Phonetics*. Second Edition. New York: Oxford University Press.

Crothers, John. 1978. Vowel Systems. In *Universals of Human Language: Volume 2-Phonology*, ed by. Joseph H. Greenberg, pp. 94-152. Stanford, CA: Stanford University Press.

Esling, John. 2013. Voice and Phonation. In *The Bloomsbury Companion to Phonetics*, ed. by Mark J. Jones and Rachael-Anne Knight, eds., pp. 110-125. New York: Bloomsbury Academic.

Esposito, Christina M. 2010. Variation in Contrastive Phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association* 40 (2): 181-198.

Fletcher, Harvey. 1940. Auditory Patterns. *Reviews of Modern Physics*12:47-65.

Gandour, Jackson T. 1978. The Perception of Tone. *Tone: A Linguistic Survey*, pp. 41-76, ed. by Victoria Fromkin. New York: Academic Press.

Gerratt, Bruce R. and Jody Kreiman. 2001. Toward a Taxonomy of Nonmodal Phonation. *Journal of Phonetics* 29:365-381.

Gordon, Matthew and Peter Ladefoged. 2001. Phonation Types: A Cross-Linguistic Overview. *Journal of Phonetics* 29: 383-406.

Hazan, Valerie. 2017. Speech Communication Across the Life Span. *Acoustics Today* 13 (1): 36-43.

Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. Acoustic Characteristics of American English Vowels. *Journal of the American Acoustical Society*, 97 (5) 3099-3111.

Keating, Patricia., Garellek, M., and Kreiman, J. 2015. Acoustic Properties of Different Kinds of Creaky Voice," in *Proceedings of the 18th International Congress of Phonetic Sciences*, the University of Glasgow, Glasgow, UK, pp. 0821.1–0821.5

Keating, Patricia and Grace Kuo. 2012. Comparison of Speaking Fundamental Frequency in English and Mandarin. *Journal of the Acoustical Society of America* 132 (2):1050-1060.

Kent, Ray D. and Charles Read. 2002. *Acoustic Analysis of Speech*. Second Edition. Clifton Park, NY: Delmar Cengage Learning.

Koffi, Ettien. 2012. Acoustic Phonetic Self-Discovery: Editor's Prologue. *Linguistic Portfolios* 1:1-4.

Koffi, Ettien and James Lyons. 2018. Sibling Acoustics: Phonetic Measurements of the Vowels Produced by Three Brothers Who "Sound Alike." *Linguistic Portfolios* 7:80-94.

Kuang, Jianjing 2017. Covariation between Voice Quality and Pitch: Revisiting the Case of Mandarin Creaky Voice. *Journal of the Acoustical Society of America* 142 (3): 1693-1706.

Ladefoged, Peter. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques.* Malden, MA: Blackwell Publishing.

Lopez-Backstrom, Michel. 2018. *An Acoustic Phonetic Analysis of Northern Minnesota English Vowel Spaces*. St. Cloud State University: MA Thesis in Applied Linguistics.

Miller, James D. 1998. Auditory-Perceptual Interpretation of the Vowel. *Journal of the Acoustical Society of America* 85(5): 2114-2134.

Ohala, John. 1978. Production of Tone. *Tone: A Linguistic Survey*, ed. by Victoria Fromkin, pp.5-39. New York: Academic Press.

Peterson, Gordon F and Harold L. Barney. 1952. Control Methods Used in a Study of the Vowels. *Journal of the American Acoustical Society*, 24 (2): 175-184.

Scharf, Bertram 1961. Complex Sounds and Critical Bands. *Psychological Bulletin*, Volume 50, 3:205-217. Retrieved from

Stevens, Kenneth. 2000. *Acoustic Phonetics*. MIT Press.

Stevens, Kenneth N. 1998. Models of Speech Production. In *Handbook of Acoustics*, ed. by Malcolm J. Crocker, pp. 1231-1244. New York: A Wiley-Interscience Publication, John Wiley and Sons, Inc.,

Titze, Ingo R. 1994. *Workshop on Acoustic Voice Analysis: Statement Summary*. Iowa City, IA: National Center for Voice and Speech.

Watt, Dominic. 2013. *Research Methods in Acoustics*. In *The Bloomsbury Companion to Phonetics*, ed. by Mark J. Jones and Rachel-Anne Knight, pp. 79-97. New York: Bloomsbury Academic.

Yost, William A. 2015. Psychoacoustics: A Brief Historical Overview. *Acoustics Today: A Publication of the Acoustical Society of America*, Volume 11 (3) 46-53.

Yost, William A. 2007. *Fundamentals of Hearing: An Introduction*. 5th Edition. New York: Elsevier

Young, Hugh D. 2011. *College Physics*. 9th edition. NY:Addison-Wesly.

Appendix

After I finished all the statistical analyses of the F0 measurements from Central Minnesota speakers, my former graduate student, Backstrom-Lopez (2018), completed her MA thesis on the acoustic phonetic characteristics of Northern Minnesota English. Included in the appendix are her F0 measurements of male and female participants in her study.

Lexical Set	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Vowels	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1M	121	109	114	108	116	105	109	126	118	131	117
Speaker 2M	174	145	133	130	136	134	129	137	135	153	129
Speaker 3M	127	132	126	127	123	127	141	133	144	140	130
Speaker 4M	112	109	100	97	91	92	85	93	99	113	100
Speaker 5M	139	131	126	119	120	120	123	N/A	131	131	N/A
Speaker 6M	100	101	110	121	N/A	105	94	109	112	124	105
Speaker 7M	127	115	123	113	121	117	121	113	126	138	103
Speaker 8M	110	115	110	110	108	108	112	116	122	128	110
Speaker 9M	120	117	118	114	112	N/A	117	126	129	144	121
Speaker 10M	101	81	100	79	79	85	76	85	84	98	83
Average	**123**	**115**	**116**	**111**	**111**	**110**	**110**	**115**	**120**	**130**	**110**
St. Deviation	21.61	17.85	11.2	14.97	17.33	15.87	20.27	17.58	17.78	15.79	15.11

F0 for Men from Northern Minnesota

Lexical Set	**fleece**	**kit**	**face**	**dress**	**trap**	**lot**	**cloth**	**goat**	**foot**	**goose**	**strut**
Vowels	[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Speaker 1F	168	146	153	147	143	156	151	166	175	183	155
Speaker 2F	220	191	204	203	192	194	199	206	217	224	201
Speaker 3F	221	204	202	191	188	188	187	197	208	214	153
Speaker 4F	198	191	193	182	166	171	186	201	204	205	226
Speaker 5F	179	197	189	184	176	187	215	202	206	219	188
Speaker 6F	223	201	181	189	183	184	196	195	193	215	186
Speaker 7F	194	191	144	170	146	161	115	148	193	189	150
Speaker 8F	286	272	257	245	241	286	199	239	252	269	136
Speaker 9F	205	185	173	170	166	176	192	N/A	175	183	168
Speaker 10F	211	196	192	194	187	159	179	193	168	187	93
Average	**211**	**197**	**188**	**187**	**178**	**186**	**181**	**194**	**199**	**208**	**165**
St. Deviation	32.28	30.8	31.06	25.6	27.69	37.49	28.79	24.02	24.65	26.32	37.24

F0 for Women from Northern Minnesota